



## آشکار سازی وضعیت تیروئید، پرولاکتین و مورفولوژی تخمدان به عنوان محورهای غالب ناهمگونی PCOS در بیماران IVF با استفاده از خوشه بندی

### کی-پروتوتایپ مبتنی بر هوش مصنوعی

آرش ضیائی<sup>۱</sup>، نیکو رضایی بروجردی<sup>۲</sup>، علیرضا شبانی<sup>۳</sup>، ملیحه محمودی نیا<sup>۴</sup>

۱. پزشک عمومی، مرکز تحقیقات بیماری‌های عفونی، دانشگاه علوم پزشکی بیرجند، بیرجند، ایران.
۲. دپارتمان پزشکی هسته‌ای، مرکز تحقیقات پزشکی هسته‌ای و تصویربرداری سلولی، بیمارستان آموزشی نمازی، دانشگاه علوم پزشکی شیراز، شیراز، ایران.
۳. دپارتمان مهندسی صنایع، دانشکده مهندسی، دانشگاه علوم پزشکی تهران، تهران، ایران.
۴. استادیار زنان و مامایی، فلوشیپ نازایی، هسته مرکزی حمایت از خانواده و جوانی جمعیت، دپارتمان زنان و مامایی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.

ایمیل نویسنده مسئول: mahmoudiniam@mums.ac.ir

### چکیده

#### مقدمه

سندرم تخمدان پلی کیستیک (PCOS) تنوع بالینی و بیولوژیکی قابل توجهی نشان می‌دهد. اگرچه معیارهای روتردام بیماران را به فنوتیپ‌های A تا D دسته بندی می‌کند، بسیاری از تظاهرات بالینی بین این برچسب‌ها قرار می‌گیرند. ما این تنوع پنهان را در زنانی که تحت درمان لقاح آزمایشگاهی (IVF) بودند، با خوشه بندی آن‌ها توسط یک الگوریتم داده‌های ترکیبی که بر پرولاکتین، عملکرد تیروئید و یافته‌های دقیق اولتراسوند تخمدان تأکید داشت، بررسی کردیم.

#### روش‌ها

زنانی که به یک مرکز ناباروری مراجعه کرده، واجد تعریف روتردام برای PCOS بودند و کاندید IVF شده بودند، در مطالعه گنجانده شدند. داده‌های بالینی استاندارد، پانل‌های بیوشیمیایی ناشتا و سونوگرافی ترانس‌واژینال ثبت گردید. مجموعه ویژگی‌ها شامل: پرولاکتین سرم، تشخیص دسته‌های کم کاری یا پرکاری تیروئید (برگرفته از هورمون محرک تیروئید (TSH) سرم)، تعداد فولیکول و حجم تخمدان چپ، به همراه سن، گنادوتروپین‌ها (هورمون محرک فولیکول و هورمون لوتئینه کننده)، نشانگرهای متابولیک (گلوکز ناشتا، کلسترول کل، تری گلیسیرید)، آنزیم‌های کبدی (آسپاراتات و آلانین آمینوترانسفراز)، ۲۵-هیدروکسی ویتامین D سرم، شاخص توده بدنی (BMI)، الگوی قاعدگی، نمره هیرسوتیسم و مورفولوژی دوطرفه تخمدان بود. الگوریتم کی-پروتوتایپ (با پنج مقداردهی اولیه، بذردهی Cao و ۱۰۰۰ تکرار) اجرا شد.

#### نتایج

الگوریتم کی-پروتوتایپ با موفقیت، ۵۱۶ زن را به پنج زیرگروه متمایز و از نظر بالینی منسجم تقسیم کرد. کیفیت خوشه بندی به طور کمی توسط چندین معیار داخلی اعتبارسنجی شد: هزینه کی-پروتوتایپ (K-Prototypes Cost) در ۱۹۱/۰۲ کمینه شد. انسجام و تفکیک خوشه‌ها با نمره سیلوئت (Silhouette Score) بالای ۰/۸۱، شاخص دیویس-بولدین (Davies-Bouldin Index) پایین ۰/۴۷ و شاخص کالینسکی-هاراباس (Calinski-Harabasz Index) بالای ۵۹۶/۰۹ تأیید شد. خوشه ۰ (۳۴۶ نفر، ۶۷٪) تصویر "معمول" را نشان داد که تخمدان‌های پلی کیستیک تقریباً دوطرفه، هیرسوتیسم متوسط و قاعدگی‌های نامنظم، همراه با شیوع بالای غیرمنتظره بیماری تیروئید (۹۹/۷٪ غیرطبیعی)؛ میانگین TSH 2.14  $\mu\text{IU/mL}$ ؛ اکثر آن‌ها با فنوتیپ A روتردام مطابقت داشتند.



خوشه ۱ (۱۶ نفر، ۳٪) با کم‌کاری آشکار تیروئید (میانگین  $TSH 6.3 \mu IU/mL$ )، هیپرتری‌گلیسیریدمی شدید ( $273 \text{ mg/dL}$ ) و بالاترین نمرات هیرسوتیسم متمایز شد، اما مورفولوژی تخمدان متناقضی داشت؛ بیماران عمدتاً به فنوتیپ‌های A و B نگاهت شدند. خوشه ۲ (۶۵ نفر، ۱۳٪) شامل زنانی با اولیگو-آنوولاسیون بارز (۷۲٪) و هیرسوتیسم (۸۰٪) اما بدون اختلال عملکرد تیروئید تیروئید (میانگین  $TSH 2.15 \mu IU/mL$ ) و اغلب با ظاهر تخمدان طبیعی یا یک‌طرفه بود؛ این گروه با فنوتیپ B همسو بود و سبک‌ترین پروفایل متابولیک را داشت. خوشه ۳ (۵۷ نفر، ۱۱٪) ترکیبی از تخمدان‌های پلی‌کیستیک دوطرفه و بیشترین اختلال قاعدگی (۷۷٪) را با ناهنجاری تیروئید همگانی (میانگین  $TSH 3.26 \mu IU/mL$ ) و افزایش خفیف ترانس‌آمینازها نشان داد که به استرس کبدی اولیه اشاره داشت؛ دو سوم آن‌ها فنوتیپ A بودند. خوشه ۴ (۳۲ نفر، ۶٪) بر هیپرپرولاکتینمی قابل توجه (میانگین  $430 \text{ ng/ml}$ ) متمرکز بود؛ اکثر زنان همچنین دارای تخمدان‌های پلی‌کیستیک و بی‌نظمی چرخه بودند، در حالی که ۸۴٪ تست‌های تیروئید طبیعی داشتند، که به این گروه یک عامل محرک اندوکرین متمایز می‌داد که فراتر از دسته‌بندی‌های روتردام بود.

### نتیجه‌گیری

خوشه‌بندی کی-پروتوتایپ، با تمرکز بر غلظت پرولاکتین، وضعیت تیروئید و سونوگرافی تخمدان، زنان مبتلا به PCOS و جویای درمان IVF را به پنج پروفایل بالینی تکرارپذیر تقسیم کرد که فراتر از طرح روتردام عمل می‌کنند. داده‌ها بر سه محور غالب ناهمگونی تأکید دارند که شامل اختلال عملکرد تیروئید، هیپرپرولاکتینمی و استرس متابولیک-کبدی است و پیشنهاد می‌کنند که ارزیابی روتین این ویژگی‌ها می‌تواند راهنمای مداخلات فردی‌سازی‌شده‌تر در سبک زندگی، غدد و متابولیسم باشد.

**کلمات کلیدی:** سندرم تخمدان پلی‌کیستیک؛ خوشه‌بندی کی-پروتوتایپ؛ اختلال عملکرد تیروئید؛ هیپرپرولاکتینمی؛ مورفولوژی تخمدان



## AI-Driven k-Prototypes Clustering Reveals Thyroid Status, Prolactin, and Ovarian Morphology as Dominant Axes of PCOS Heterogeneity in IVF Patients

Arash Ziaee<sup>1</sup>, Nikoo Rezaei Borujerdi<sup>2</sup>, Alireza Shabani<sup>3</sup>, Maliheh Mahmoudinia<sup>4,\*</sup>

1. General Physician, Infectious Diseases Research Center, Birjand University of Medical Sciences, Birjand, Iran.
2. Department Nuclear Medicine, Nuclear Medicine and Molecular Imaging Research Center, Namazi Teaching Hospital, Shiraz University of Medical Sciences, Shiraz, Iran.
3. Department of Industrial Engineering, Faculty of Engineering, College of Farabi, University of Tehran, Tehran, Iran.
4. Assistant Professor of Obstetrics & Gynecology, Fellowship of Infertility, Supporting the Family and the Youth of Population Research Core, Department of Obstetrics and Gynecology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

\* Corresponding author e-mail: [mahmoudiniam@mums.ac.ir](mailto:mahmoudiniam@mums.ac.ir)

### Abstract

#### Introduction

Polycystic ovary syndrome (PCOS) shows marked clinical and biological variety. Although the Rotterdam criteria organize patients into phenotypes A–D, many presentations fall between these labels. We explored this hidden diversity in women undergoing in-vitro fertilization by clustering them with a mixed-data algorithm that emphasized prolactin, thyroid function and detailed ovarian-ultrasound findings.

#### Methods

Women attending a single infertility center who met the Rotterdam definition of PCOS and were scheduled for IVF were included. Standard clinical data, fasting biochemical panels and transvaginal sonography were recorded. The feature set consisted of serum prolactin, categorical diagnosis of hypo- or hyper-thyroidism derived from serum thyroid-stimulating hormone (TSH), and left-ovary follicle number and volume, alongside age, gonadotrophins (follicle-stimulating hormone and luteinizing hormone), metabolic markers (fasting glucose, total cholesterol, triglycerides), liver enzymes (aspartate and alanine aminotransferases), serum 25-hydroxy-vitamin D, body-mass index, menstrual pattern, hirsutism score and bilateral ovarian morphology. A k-prototypes algorithm (five initializations, Cao seeding,  $\gamma=2.178$ , 1000 iterations) was run with  $k = 5$ .

#### Results

The k-prototypes algorithm successfully separated the 516 women into five distinct and clinically coherent subgroups. The quality of the clustering was quantitatively validated by several internal metrics: the K-Prototypes Cost was minimized at 191.02. Cluster cohesion and separation were further supported by a high Silhouette Score of 0.81, a low Davies-Bouldin Index of 0.47, and a high Calinski-Harabasz Index of 596.09. Cluster 0 ( $n = 346$ , 67 %) represented the “typical” picture: almost universal bilateral polycystic ovaries, moderate hirsutism and infrequent menses, together with an unexpectedly high burden of thyroid disease (99.7 % abnormal; mean TSH 2.14  $\mu\text{IU/mL}$ ); most matched Rotterdam phenotype A. Cluster 1 ( $n = 16$ , 3 %) was distinguished by overt hypothyroidism (mean TSH 6.3  $\mu\text{IU/mL}$ ), severe



hypertriglyceridemia (273 mg/dL) and the highest hirsutism scores, but showed inconsistent ovarian morphology; patients mapped chiefly to phenotypes A and B. Cluster 2 (n = 65, 13 %) consisted of women with pronounced oligo-anovulation (72 %) and hirsutism (80 %) but no thyroid dysfunction (mean TSH 2.15  $\mu$ IU/mL) and often normal or unilateral ovarian appearance; it aligned with phenotype B and had the lightest metabolic profile. Cluster 3 (n = 57, 11 %) combined bilateral polycystic ovaries and the greatest menstrual disturbance (77 %) with universal thyroid abnormality (mean TSH 3.26  $\mu$ IU/mL) and mild transaminase elevation, hinting at early hepatic stress; two-thirds were phenotype A. Cluster 4 (n = 32, 6 %) centered on marked hyperprolactinemia (mean 430 ng/mL); most women also had polycystic ovaries and cycle irregularity, while 84 % showed normal thyroid tests, giving this group a distinct endocrine driver that cut across Rotterdam categories.

### Conclusion

By concentrating on prolactin concentration, thyroid status and ovarian ultrasonography, k-prototypes clustering divided IVF-seeking women with PCOS into five reproducible clinical profiles that transcend the Rotterdam scheme. The data emphasize three dominant axes of heterogeneity—thyroid dysfunction, hyperprolactinemia and metabolic-hepatic stress—and suggest that routine evaluation of these features could guide more individualized lifestyle, endocrine and metabolic interventions.

**Keywords:** Polycystic Ovary Syndrome; k-Prototypes Clustering; Thyroid Dysfunction; Hyperprolactinaemia; Ovarian Morphology